DOCUMENT RESUME

ED 209 324 TM 810 864

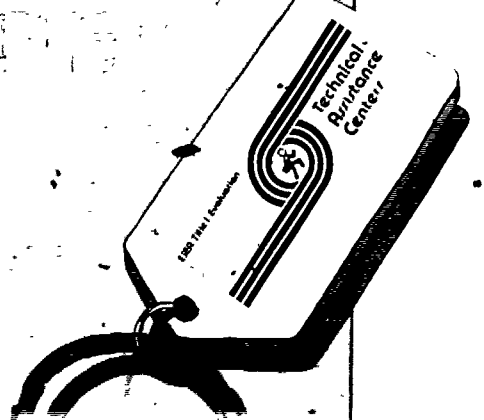
AUTHOR Demaline, Randy; Rader, Don

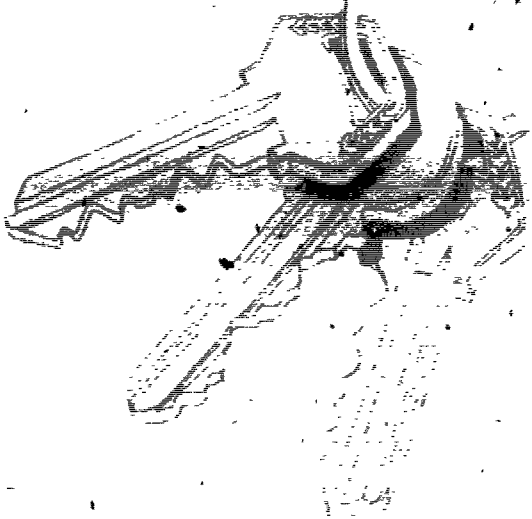
TITLE Model A Owner's Manual.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY Department of Education, Washington, D.C.
PUB DATE 81
CONTRACT 300-79-0-488; 300-79-0-489; 300-79-0-490
NOTE 27p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Methods; *Models; Program Effectiveness; *Program Evaluation
IDENTIFIERS *Elementary Secondary Education Act Title I; *RMC Models; Title I Evaluation and Reporting System

ABSTRACT

This manual briefly reviews the major guidelines for a Title I evaluation plan using the norm-referenced evaluation model, Model A-1. Designed to help local school district personnel responsible for Title I evaluation activities, the manual can be used as a quick check of present evaluation activities or as a planning guide for future activities. The manual is divided into three sections. In the first, Title I Model A Testing Plan, a form is provided for recording significant information about a testing plan. In the second section, Title I Evaluation Guidelines, a series of clipped pages with arrows at the edges refer the reader to certain sections of the testing form. The pages provide information about the following guidelines: how using Model A measures the impact of Title I programs, separation of pretest and selection, test forms and levels, testing at empirical norm dates, and measuring sustained effects. Additional information is presented in the third section. This refers to ideas mentioned in the guidelines which provide more information on concepts involved in Title I evaluation. This section includes normal curve equivalents, why selection and pretest should be separate, functional-level testing, why it's important to test at norm dates, and Title I evaluation technical standards. (Author/GK)





MODEL A
OWNER'S MANUAL

Developed by:

Randy Demaline
Don Rader

Distributed by:

Title I Technical Assistance Centers
Regions 8, 9 and 10
Northwest Regional Educational Laboratory
710 S.W. Second Avenue
Portland, Oregon 97204

Purpose of Owner's Manual

This manual briefly reviews the major guidelines for a Title I evaluation plan using the norm-referenced evaluation model, Model A-1. Specifics about Model A-2, using a non-normed test, are not covered in this manual.

Designed to help local school district personnel responsible for Title I evaluation activities, the manual can be used as a quick check of present evaluation activities or as a planning guide for future activities.

This project has been funded with Federal funds from the United States Department of Education under Contract numbers 300 790 488, 300 790 489 and 300 790 490. The contents do not necessarily reflect the views or policies of the Department of Education, nor does mention of trade names, commercial products or organizations imply endorsement by the U.S. Government.

Description of Contents

This manual is divided into three sections:

Title I Model A Testing Plan

A form is provided on page iv for recording significant information about your present or anticipated testing plan. If you want to use the form for planning the evaluation, you should duplicate it since a separate form is needed for each project and for different grades within each project.

Title I Evaluation Guidelines

A series of clipped pages follow the Testing Plan form. Arrows at the edges of these pages refer you to certain sections of your testing form. The pages provide information about the following guidelines:

Additional Information

The rest of the manual contains sections providing additional information about some of the ideas mentioned in the guidelines. They may provide a more in-depth rationale for the guidelines or more information on concepts involved in Title I evaluation. They are not meant to be technical explanations nor detailed steps in the implementation. If you need more information, contact your Service Center identified on page 21.

Title I Model A Testing Plan

for _____ _____
project name grade

The following information should be completed for each project and for each grade level within a project.

Student Selection Information

List any subtest names that will be used in selecting students for Title I instruction (Selection of eligible students for Title I instruction may involve more than one test and other non-test data that have been identified as useful.)

_____ from _____
subtest name name of test

_____ Date administered to students

Pretest Information

_____ Subtest name from _____
subtest name name of test battery

_____ Date administered to students

_____ Empirical norm date (*see test battery manual*)

_____ Level(s) of the test you are using

_____ Form of test you are using

Posttest Information

_____ Subtest name from _____
subtest name name of test battery

_____ Date administered to students

_____ Empirical norm date

_____ Level(s) of the test you are using

_____ Form of test you are using

Sustained Effects Information

_____ Subtest name from _____
subtest name name of test battery

_____ Date (*month and year*) of administration

Title I Evaluation Guidelines

HOW USING MODEL A MEASURES THE IMPACT OF YOUR TITLE I PROGRAM

To Whom Are Title I Students Compared?

Your Title I students are compared to a national sample of students who score at the same pretest percentile. For example, if your Title I class scores at the 25th percentile at the pretest, its growth will be compared to the students in the norm group who scored at the 25th percentile.

Estimating How the Students Would Have Done Without Title I Help

Model A's assumption is that a group of students who do not receive Title I instruction will maintain a constant percentile throughout a year. For example, a group of *non*-Title I students who scored at the 25th percentile at the beginning of the year is expected to maintain the same relative standing (25th percentile) throughout the year. The achievement level of these students would increase but so would the achievement levels of students above and below the 25th percentile).

The Definition of "Title I Impact"

Any change in the percentile rank of a group of Title I students from the pretest to the posttest is attributed to Title I instruction. Thus, if Title I students scored at the 25th percentile at the pretest and at the 30th percentile at the posttest, the increase is considered to be the Title I impact. If the Title I students score at the same pretest and posttest percentile, the program would be considered to have had no impact. students who received only regular instruction with no special help are expected to maintain the same percentile.

SEPARATION OF PRETEST AND SELECTION

Guideline

The score used for selection should *not* be used as the pretest score.

Purpose of the Guideline

If the same test score is used for both selecting Title I students and as a pretest score, you will *over*estimate your project's impact. That is, your Title I program will appear to be more effective than it really was.

Alternatives for Separating Pretest and Selection

1. Administer a separate pretest and selection test.
2. Use last year's posttest scores as this year's selection scores.
3. Use different subtests of the same test battery; one subtest for selection and the other for evaluation. Take care that the two tests are related and both should be related to the objectives of the project.
4. *Re*administer the same test for the pretest as was used for selection.

You should evaluate your selection procedures at the end of each year to make any adjustments or changes.

Exception to the Guideline

The pretest score can be used for selection only if:

1. *Selection is based totally on the pretest score such that the students assigned to the Title I program are those who score lowest on the pretest, and*
2. *the correction formula in the* User's Guide *is used to adjust (i.e. correct) the group's pretest mean. The adjusted pretest mean is then used to estimate the expected posttest percentile rank of the group.*

NOTES

For more information on separation of pretest and selection

turn to page 11

TEST FORMS AND LEVELS

Definitions

Different LEVELS of a test contain items of different difficulty.

Different FORMS of a test will be at the same level of difficulty, but will contain different but comparable items.

Levels Guideline

Administer test levels whose difficulty and content match the performance levels of the students (functional-level testing).

Forms Guideline

Use the same test forms for pretest and post-test that the publisher used with the norming group.

NOTES

For more information about functional-level testing

turn to page 14

TESTING AT EMPIRICAL NORM DATES

Guideline

Testing should occur within *two weeks* before or after the publisher's empirical norm dates.

Definition

Empirical norm dates are the dates that the test publisher *actually* administered the test to a national sample. This is as opposed to *projected norms*, where the norms are merely estimates.

Purpose of the Guideline

The norm group is the comparison group for Model A. To make accurate comparisons, the Title I group should be tested at the same time of the school year as the norm group.

Deviations from the norm dates should be in the same direction and magnitude for both pretest and posttest. For example, if pretesting occurred one week before the norm date, then posttesting should also occur one week before the norm date.

Exceptions to the Guideline

It is allowable to deviate up to six weeks either side of the norm dates. If you decide to do so, you will need to perform extra computations to adjust the evaluation results to validly reflect your project's impact.

12

MEASURING SUSTAINED EFFECTS

Measuring for sustained effects is not an official component of Model A. It is a significant and required part of your evaluation plan and so has been included in this manual.

Guideline

A followup measure for sustained effects must be administered after the posttest and at least twelve months after the pretest.

Purpose of Sustained Effects Information

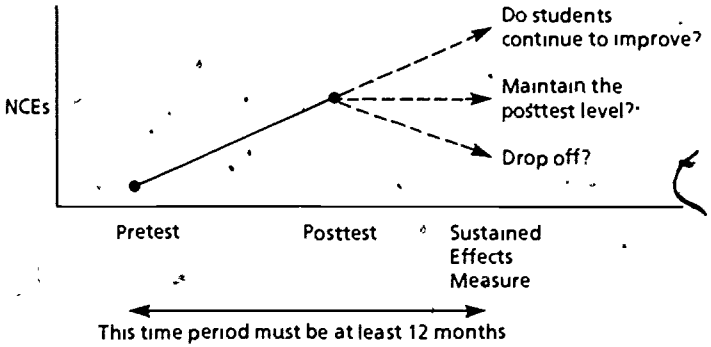
It is important that gains made by students in Title I programs from pretest to posttest be sustained over longer periods of time. Thus, a followup measure after the initial pretest-posttest cycle is needed. How soon after the posttest the information should be collected depends on what questions you want to examine (see page 9). This sustained effects information must be used in subsequent program planning.

What to Use to Measure Sustained Effects

The instrument used to measure sustained effects must be an objective measure of educational achievement in the basic skills and in the same content area as the project.

DESIGNING YOUR SUSTAINED EFFECTS STUDY

Not everyone's sustained effects study will be the same. Each project may look at different students at different times to see what happens to students after the initial pretest-posttest results reported in Model A.



Three possible questions you might ask are:

1. Are achievement gains occurring during the school year maintained *over the summer months*?
2. Do the effects of Title I instruction continue *after the students leave the program*?
3. How are students performing who were in Title I last year and *continued in the Title I program*?

Notice that these three questions differ on:

Whom to Measure—Students still in the Title I project or students who are no longer receiving services.

When to Measure—Question 1 would imply a spring posttest and a fall sustained effects measure whereas Questions 2 and 3 imply the collection of data the following spring.

Piggybacking Sustained Effects Measure With Model A Testing

Collection of sustained effects information does not always require additional testing. For example, a pretest for one year could also serve as a measure of sustained effects for a previous year.

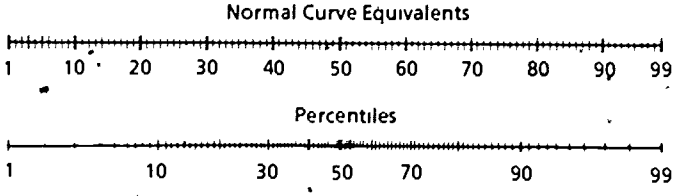
Additional Information

NORMAL CURVE EQUIVALENTS (NCEs)

You might find it easy to use percentile ranks to interpret the gains made by a group of Title I students. For example, your Title I class may have progressed from the 15th percentile at the pretest to the 18th percentile at the posttest.

However, there are problems with using percentile gains. They cannot be used for comparing your Title I group's gain to other Title I groups. They also cannot be used for combining project gains across projects (as in a district report). Percentile gains have different meanings at different places on the percentile scale. In the figure below, you will see that a gain of 10 percentiles at the high or low ends of the scale is much larger than a 10 percentile gain in the middle. (See the shaded bars.)

To avoid the problems with percentiles, NCEs were constructed to have equal units along a scale. Percentiles and NCEs match at the 1st, 50th, and 99th points on the scales. The two scales laid adjacent to one another look like this:



You can see that a gain of 10 NCEs is the same anywhere along the scale. So by using NCEs, you can more accurately compare the gains of your Title I students to those in other Title I projects.

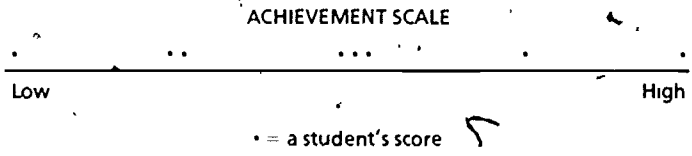
An Abbreviated Percentile to NCE Table

%ile	NCE	%ile	NCE	%ile	NCE
1	1 0	35	41 9	70	61 0
5	15 4	40	44 7	75	64 2
10	23 0	45	47 4	80	67.7
15	28 2	50	50.0	85	71.8
20	32 3	55	52 6	90	77 0
25	35 8	60	55 3	95	84.6
30	39 0	65	58.1	99	99 0

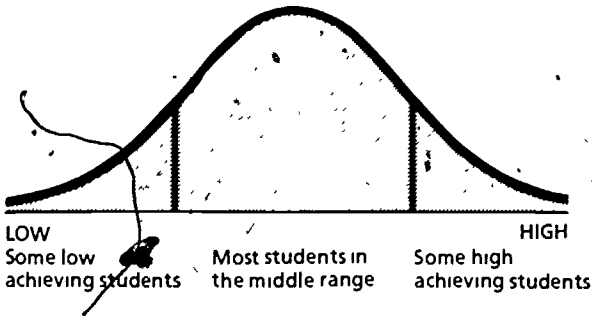
WHY SHOULD THE SELECTION AND PRETEST BE SEPARATE?

We have stated that the selection and pretest should be separate. We also have said that failure to separate them will result in an overestimation of project impact. The following explanation shows how this happens.

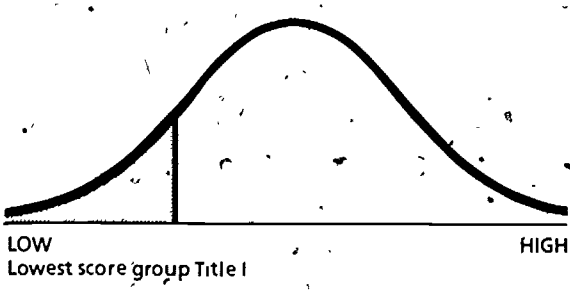
Test scores rank students on an achievement scale.



If you gave a test to all the students in a grade, you might expect to find results similar to the figure below — most of the students would score around the middle with a few students at the high and low ends of the scale.

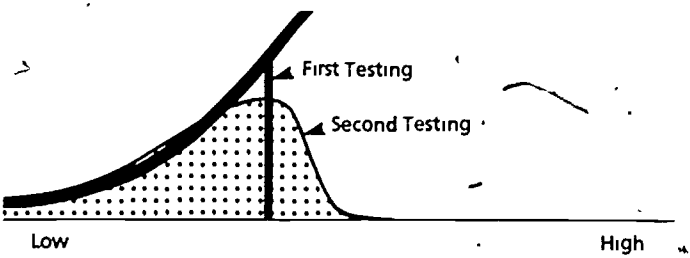


You may use a test like this to select students for Title I instruction.



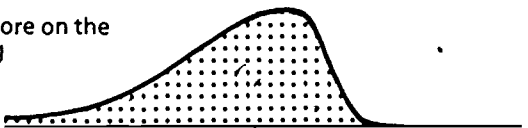
If you gave the same test to these Title I students the next day, would you expect the students to score the same?

Probably not. Some students would score higher, some, lower. The resulting curve from the second testing would be more spread out than the curve from the first testing.



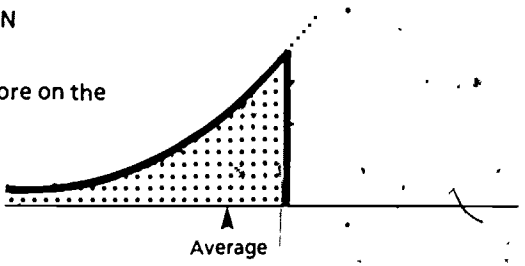
What has happened to the average score of the group between the two testings?

The average score on the Second Testing



is HIGHER THAN

the average score on the First Testing

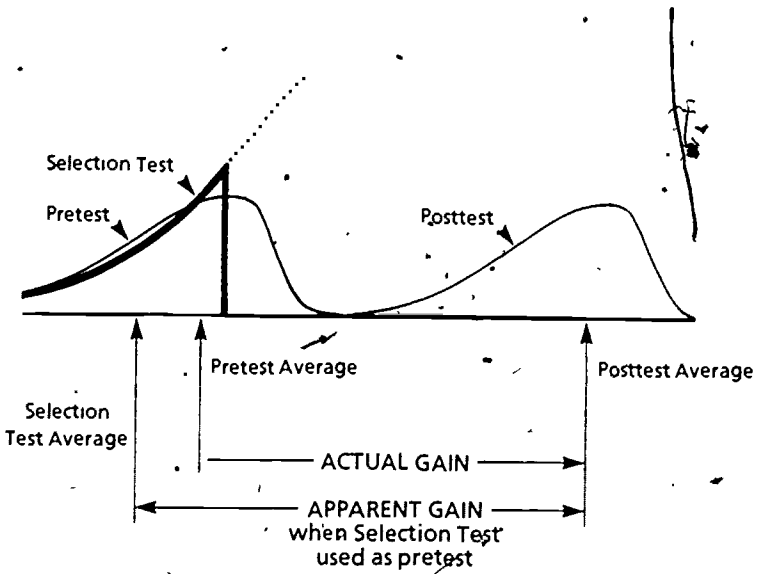


But this average is higher simply because of measurement error.

How does this artificial gain (technically known as regression to the mean) affect Title I evaluation results?

In Title I evaluation, a group's achievement gain is found by comparing its pretest average with its posttest average.

If you used the group's selection average as its pretest average, you would be including this artificial gain in the estimate of the impact of your Title I project.



The *actual gain* shown by a group is the gain from the pretest average to the posttest average.

The *apparent gain* would be the gain from the selection test average to the posttest average. It is a combination of the *actual* gain *plus* the *artificial* gain. You can see that it is an overestimation of the project's impact.

FUNCTIONAL-LEVEL TESTING

What Is It?

Giving a student a test level that has been recommended by the publisher for the student's grade level is called *in-level testing.* However, the recommended level of a test does not always contain the most appropriate content or difficulty for students with very low or very high performance levels. When testing such students, you may want to administer a test level other than the specific level recommended by the test publisher for typical students in that grade. Testing with an easier or more difficult level than recommended by the publisher is called *out-of-level* testing. Whenever you give a test level that you feel is most appropriate for a student's performance level, it is considered *functional-level testing.*

Why Do It?

Most students' functional levels are best served when the test publisher's recommended level is administered. This will usually result in a valid measure of the achievement of a group of students. However, a test that is too difficult or too easy may provide very little information about students' actual achievement. Students who are frustrated by a test that is too difficult may give up early, or they may simply guess their way through the test. If a test is too easy, students frequently lose interest. In either case, the test scores will not provide a valid or reliable assessment of achievement.

Testing at the functional level is especially important in Title I evaluation because most Title I children are low achievers. They might be tested more effectively with a lower level than recommended for most students in their grade. Obtaining valid measures of the group's ability both at the pretest and posttest will provide a more valid measure of the Title I project's impact.

What You Need for Out-of-Level Testing

If you are planning to use out-of-level testing, your test must meet some necessary criteria.

1. *The Test Must Have an Expanded Scale Score.* To be able to compare students taking different test levels, test publishers provide a way of placing all students on a common scale regardless of what level they were administered. The score from this scale is generally known as an *expanded scale score.* The exact name may differ from publisher to publisher.

   If you are considering out-of-level testing, look in your test manual and find the name of the expanded scale score. Write it here:

   _____

   Expanded Scale Score Name

2. *You Must Be Able to Follow the Subtest Used Across the Levels.* The subtest used for evaluation on the test level administered must also appear on the recommended test level. For example, if the "Word Analysis" score is used on the out-of-level test, it must also appear on the in-level test. Check:

   What test level are you going to administer? _____

   What subtest score are you using from this test? _____

   What test level does the publisher recommend? _____

   Does the subtest score occur on both the recommended and the administered levels? If not, this score canot be used for out-of-level testing.

3. *Norms Must Exist for the Students' Grade Level.* To be able to use the test in Title I Model A evaluation, norms must exist for the grade level for the students. That is, even though you may be giving 8th graders a level designed for 6th graders, norms still must be available for 8th graders.

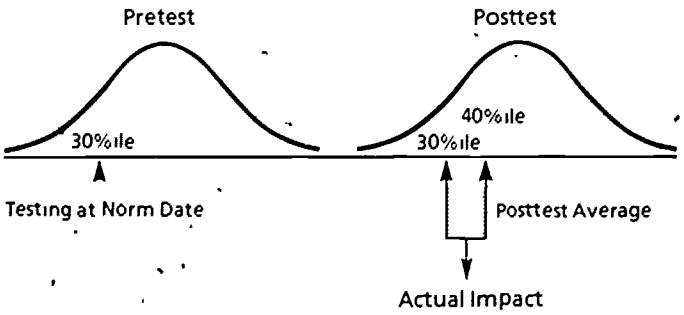
WHY IT'S IMPORTANT TO TEST AT NORM DATES

Testing at empirical norm dates is important if you expect to obtain usable data from your Model A evaluation. Let's consider the cases where the pretesting occurs at, before, or after the norm date.

Case 1: Pretesting at the Empirical Norm Date.

If you test at the norm date, your Title I students will have received approximately the same amount of instruction as the norm group did before being pretested. Your Title I group's performance on the pretest is compared to the norm group's performance and is reflected in the percentile rank. Any percentile increase at the posttest time is the program's impact.

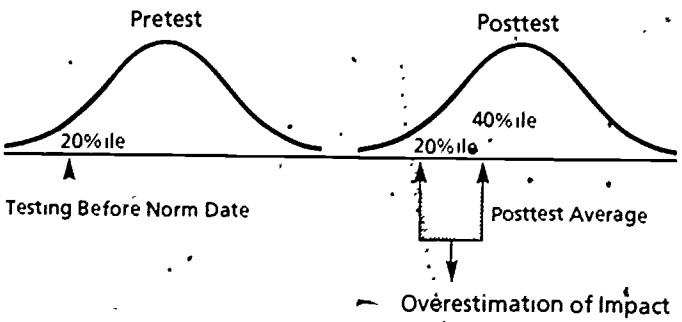
Example: Suppose you pretested a Title I group of students at the empirical norm date and found that as a group they scored at the 30th percentile. If at the posttest the group scored at the 40th percentile, you would be correct in stating that your Title I program increased the group's standing from the 30th to the 40th percentile.



Case 2: Pretesting Before the Norm Date

If you pretested your students too soon before the norm date, your students would have considerably *less* instruction than students in the norm group had before they were pretested. They would not be expected to do as well as the norm group and they would score at a *lower* percentile.

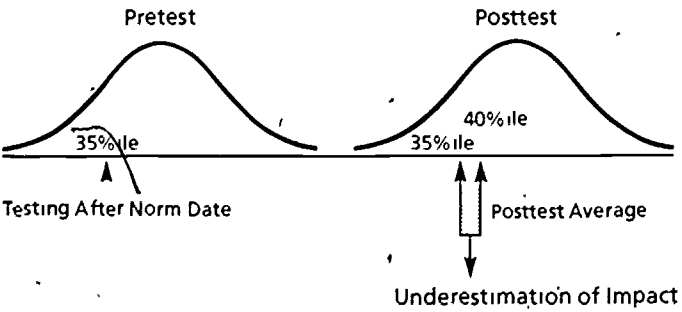
Example: Using the same group of students as in Case 1, if you pretested considerably before the norm date, they would score lower than the 30th percentile, say the 20th percentile. If you compared that with the posttest average of the 40th percentile, you would be *over*estimating the impact of the Title I program by saying it increased the group's standing from the 20th to the 40th percentile.



Case 3: Pretesting After the Norm Date

If you tested your students too far *after* the norm date, your students would have had considerably *more* instruction than students in the norm group had before they were pretested. They would probably score much *higher* than if tested at the norm date.

Example: If you pretested the students in Case 1 considerably after the norm date, they would score higher than the 30th percentile, say the 35th percentile. If you compared that with the posttest average of the 40th percentile, you would be *under*estimating the impact of the Title I program.



The results would be even more complicated if the posttesting did not occur at the norm date.

TITLE I EVALUATION TECHNICAL STANDARDS

For your Model A evaluation to run well you must follow certain maintenance requirements. In Title I evaluation, these requirements are known as the Technical Standards.

1. *Representativeness of evaluation findings.* The evaluation results must be computed so that the conclusions apply to the persons or schools served by the Title I project. This may be accomplished by including in the evaluation either all or a representative sample of the persons or schools served by the project.

2. *Reliability and validity of evaluation instruments and procedures.* The proposed evaluation instruments:
   a. Must consistently and accurately measure the objectives of the project; and,
   b. Must be appropriate, considering factors such as age or background or the persons served by the project.

3. *Evaluation procedures that minimize error.* The proposed evaluation procedures minimize error by including:
   a. Proper administration of the evaluation instruments;
   b. Accurate scoring and transcription of data; and
   c. Use of analysis procedures whose assumptions are appropriate for the data.

4. *Valid assessment of achievement gains in reading, language arts and mathematics.* In assessing the effectiveness of regular school year Title I reading, language arts, and mathematics projects in grades 2 through 12, the proposed evaluation procedures yield a valid measure of (1) the Title I children's performance after receiving Title I services compared to (2) an estimate of what their performance would have been in the absence of Title I services.

If you need more information about these standards, contact your Service Centers listed on page 21.

A WARNING FROM THE MODEL A ENGINEERS

Good drivers know the limitations of their automobiles as good evaluators know the limitations of their evaluation designs. The information gained from a Model A evaluation is limited in that it consists only of achievement data. Overinterpreting these results without considering other special programs or data such as attitudes, different achievement measures, and costs (dollars and time) would be inappropriate.

SERVICE CENTERS

For further assistance with your Model A evaluation, contact your authorized service representatives:

State Contact Name

Address

Phone Number

Northwest Regional Educational Laboratory
Title I Evaluation Technical Assistance Centers
710 S.W. Second Avenue
Portland, Oregon 97204
(503) 248-6971